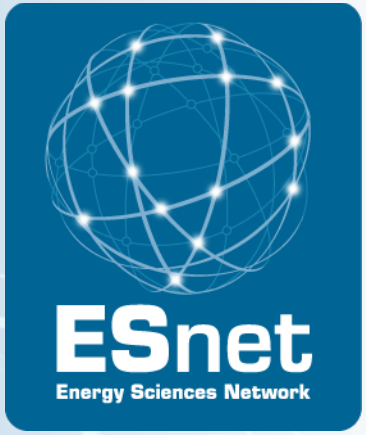


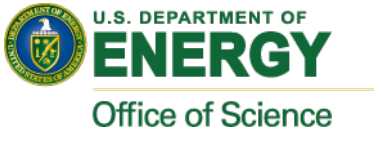
ESnet and ANI Testbed Update

Brian Tierney

DOE PI Meeting

Bethesda, MD

March 2, 2011



Outline

1. ESnet Overview
2. ESnet R&D projects: OSCARS and perfSONAR
3. ANI Testbeds

ESnet is a **Mission Organization**

DOE Office of Science mission: to deliver scientific discoveries and major scientific tools for transforming our understanding of nature, and to advance the energy security, economic security, and national security of the United States.

ESnet mission: to accelerate scientific discovery for the DOE Office of Science by delivering unparalleled network infrastructure, services, tools, and innovation.

ESnet **Supports DOE Office of Science**

SC provides broad support for Labs, Facilities, people

- almost $5B/year in funding
- 45% of Federal support for physical sciences research
- key funding for basic research in biology, computing, energy, climate
- over 100 Nobel Prizes in past 60 years

Supporting > 27,000 PhDs, grad students, engineers at >300 institutions.

Provides world's largest collection (32) of scientific user facilities:

- supercomputer centers, accelerators, light sources, neutron sources, electron microscopes, nano-scale centers, a sequencing center, fusion facilities.

ESnet is one of them – connecting DOE sites, facilities, scientists and collaborators.

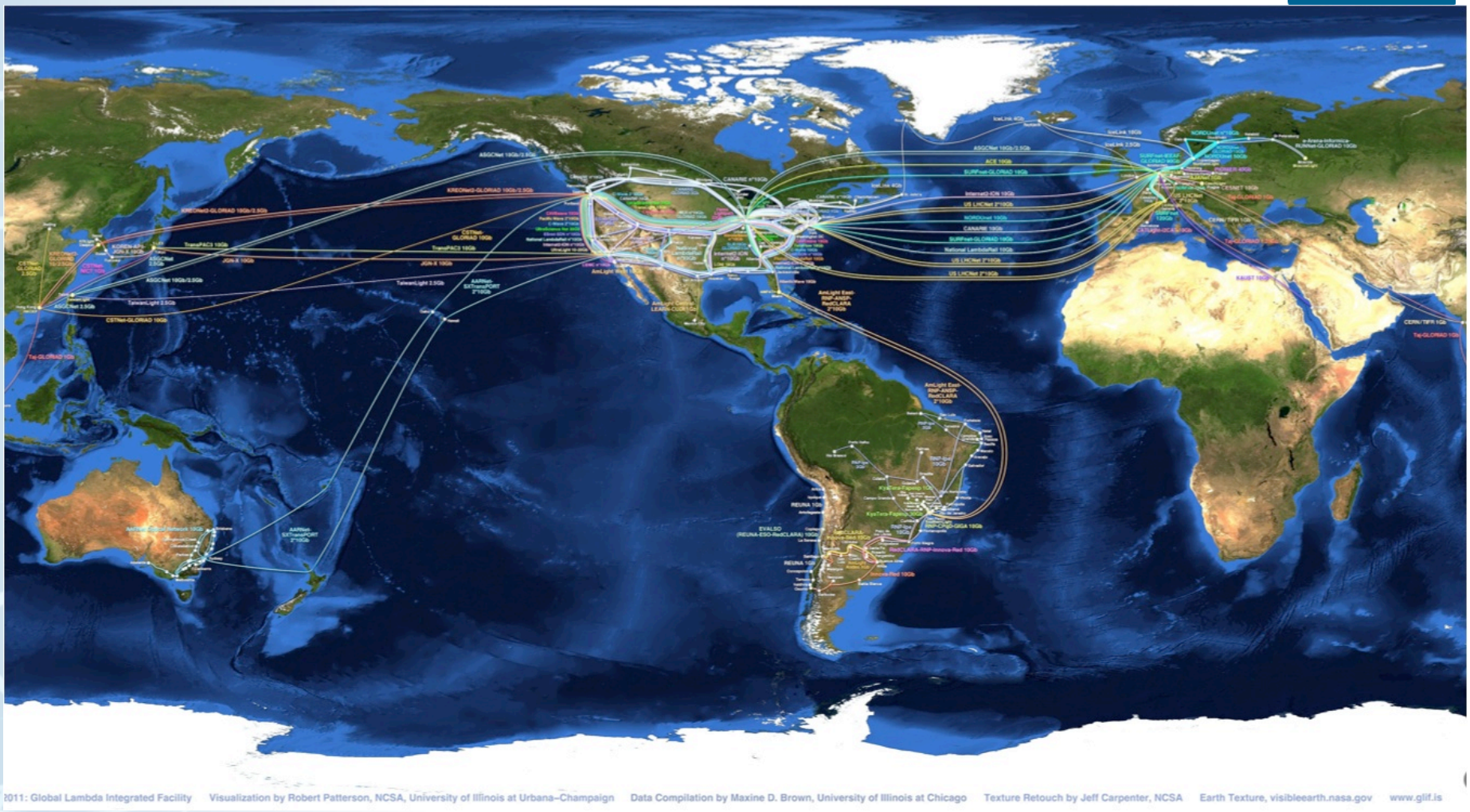
- optimized for science data transport
- every service exists to support scientific discovery

ESnet **Supports DOE Office of Science**

The Office of Science supports:

- 27,000 Ph.D.s, graduate students, undergraduates, engineers, and technicians
- 26,000 users of open-access facilities
- 300 leading academic institutions
- 17 DOE laboratories

ESnet is Embedded in a **Global Research Networking Ecosystem**



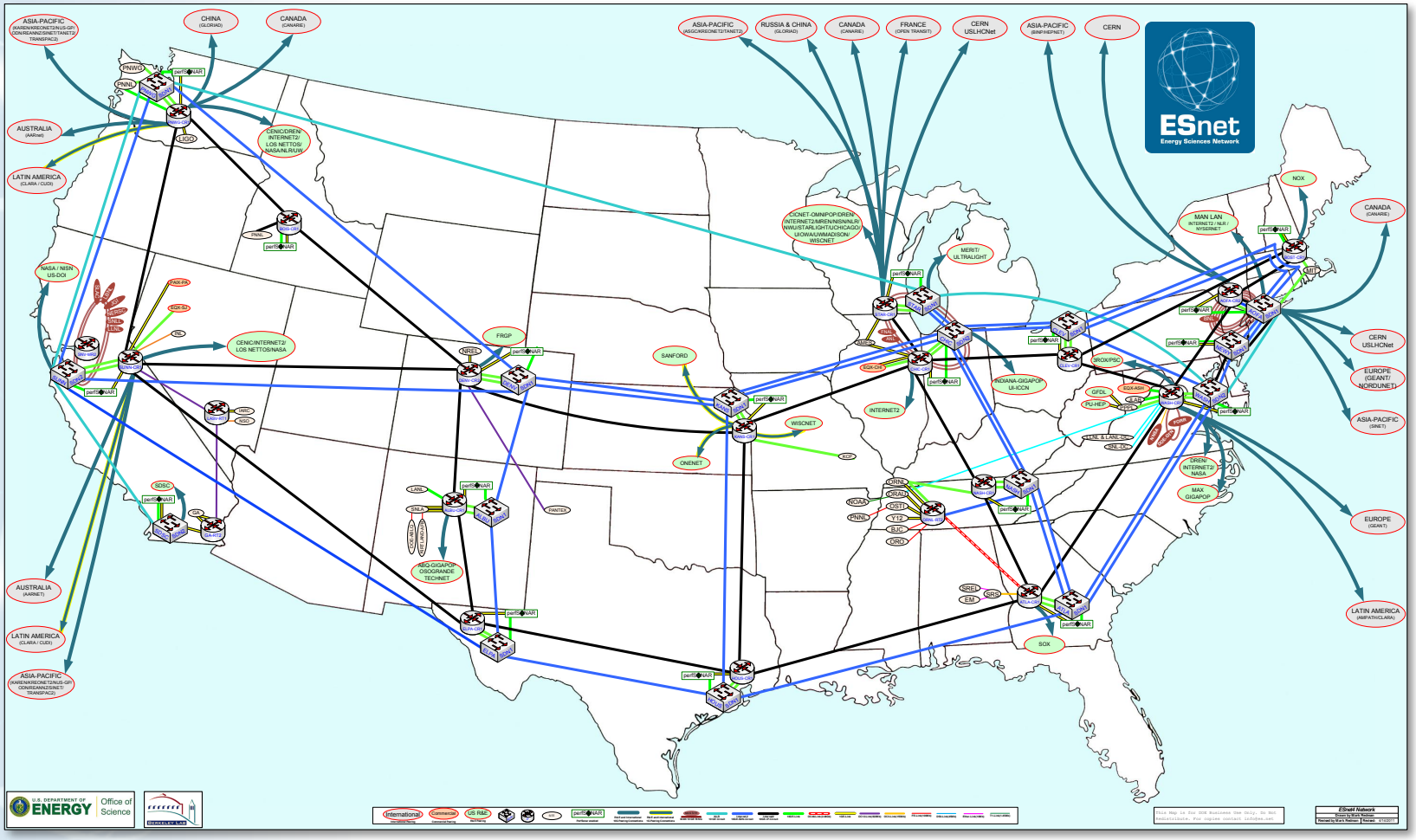
ESnet is Embedded in a **Global Research Networking Ecosystem**

Most important implication: requirement for *multi-domain* coordination.

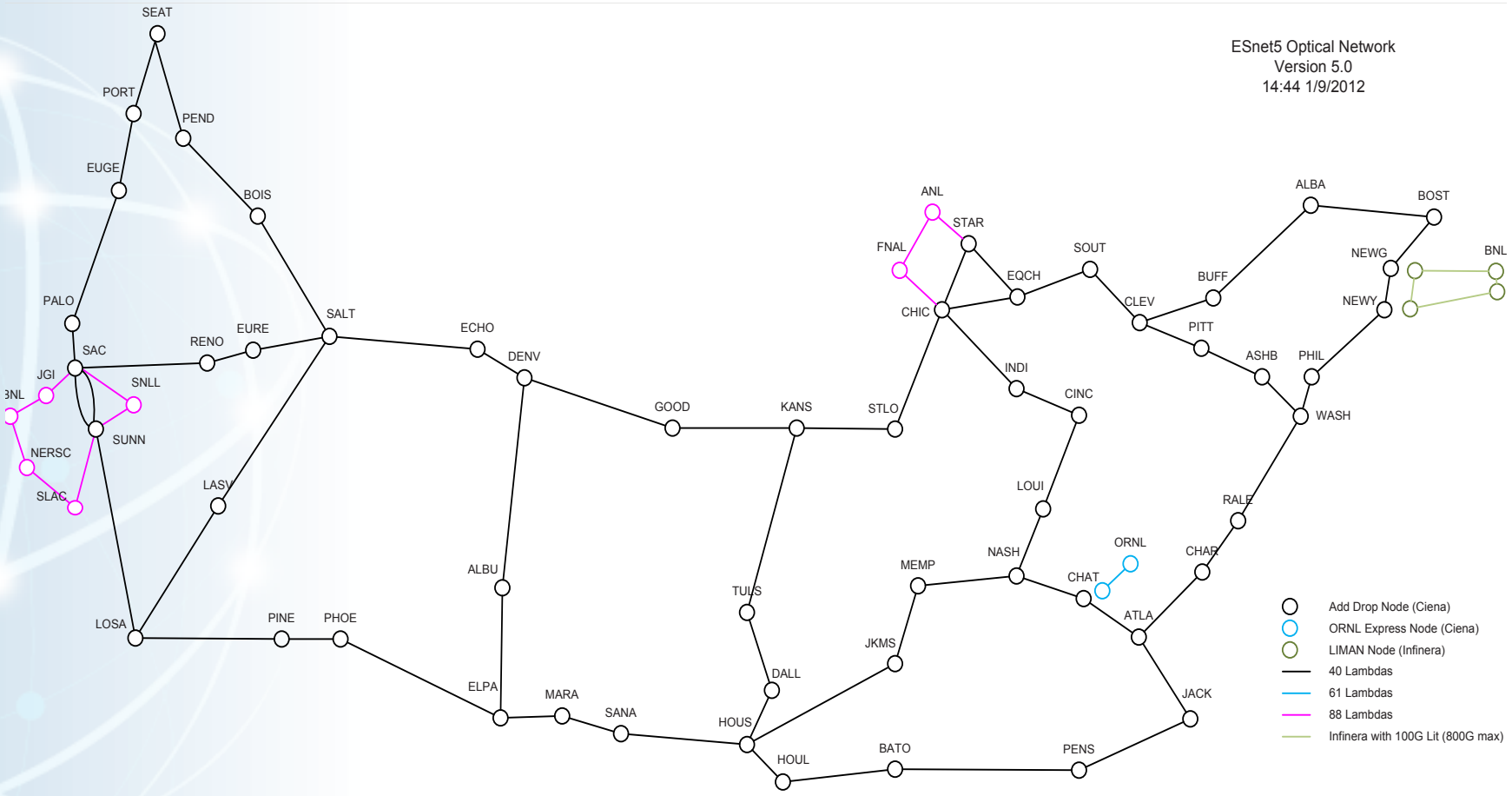
- Large-scale science is a multi-domain, end-to-end endeavor.
- We must build services across networks, exchanges, campuses.
  - Problem resolution, data mobility, measurement / monitoring, guaranteed services, collaboration – all multi-domain!
- We spend much time coordinating, communicating
  - International: GLIF, DICE, LHCOPN, LHCONE, TERENA
  - US: Joint Techs, ESCC, Supercomputing, many more

Our drivers are different from those of commercial providers.

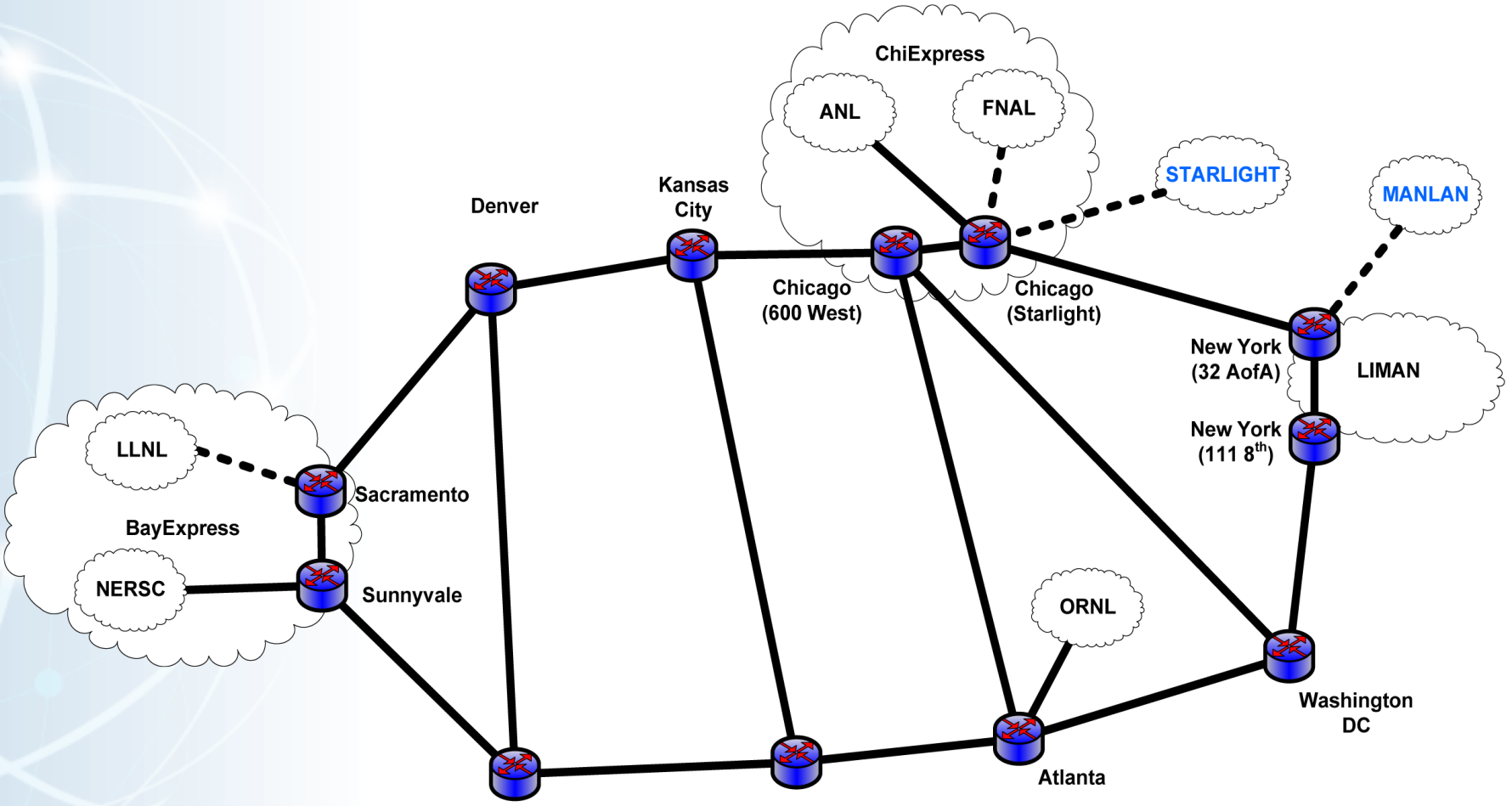
ESnet4 Topology (n x 10G core)



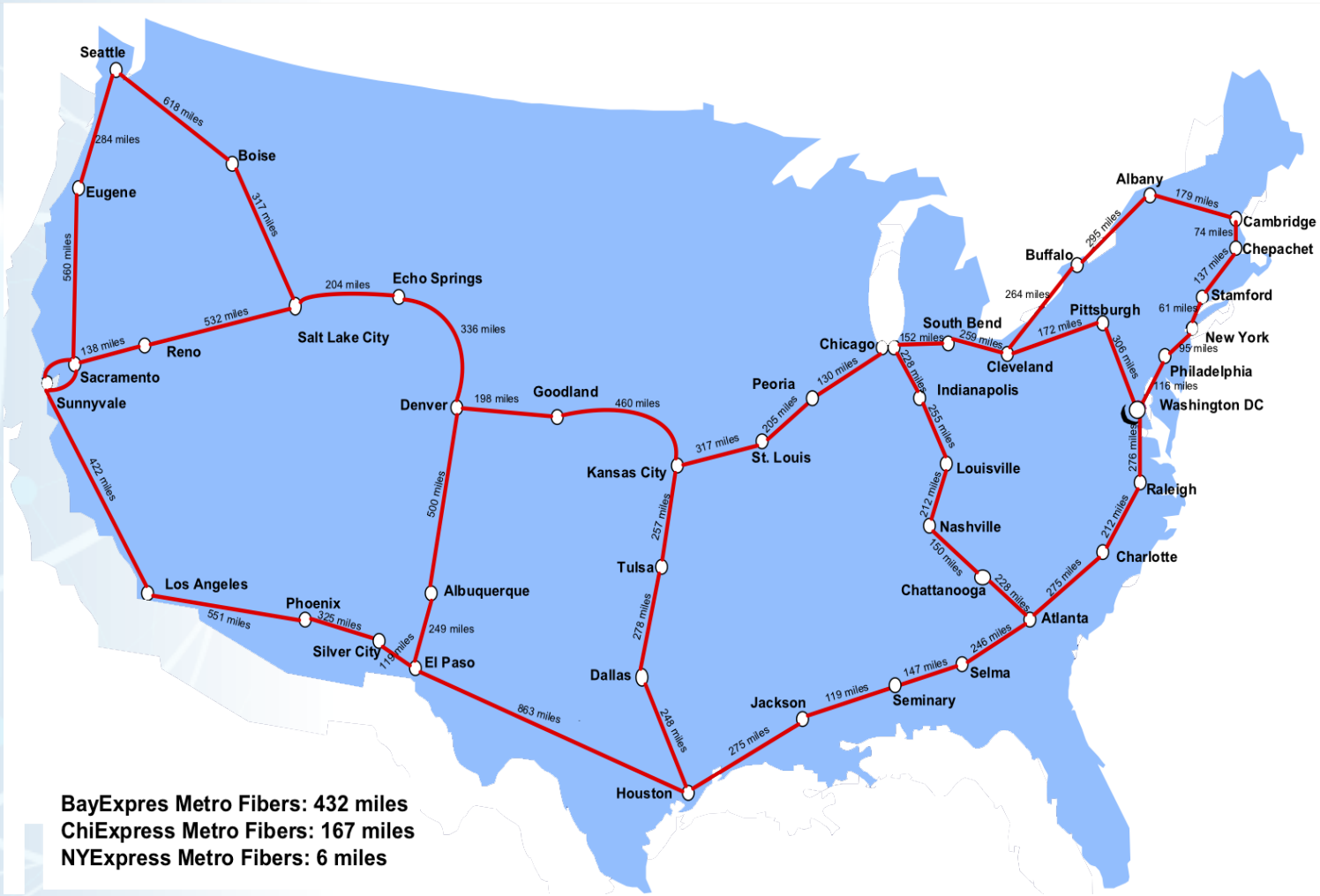
ESnet5 Optical Infrastructure Footprint



ESnet5 L2/3 Architecture (n x 100G core)

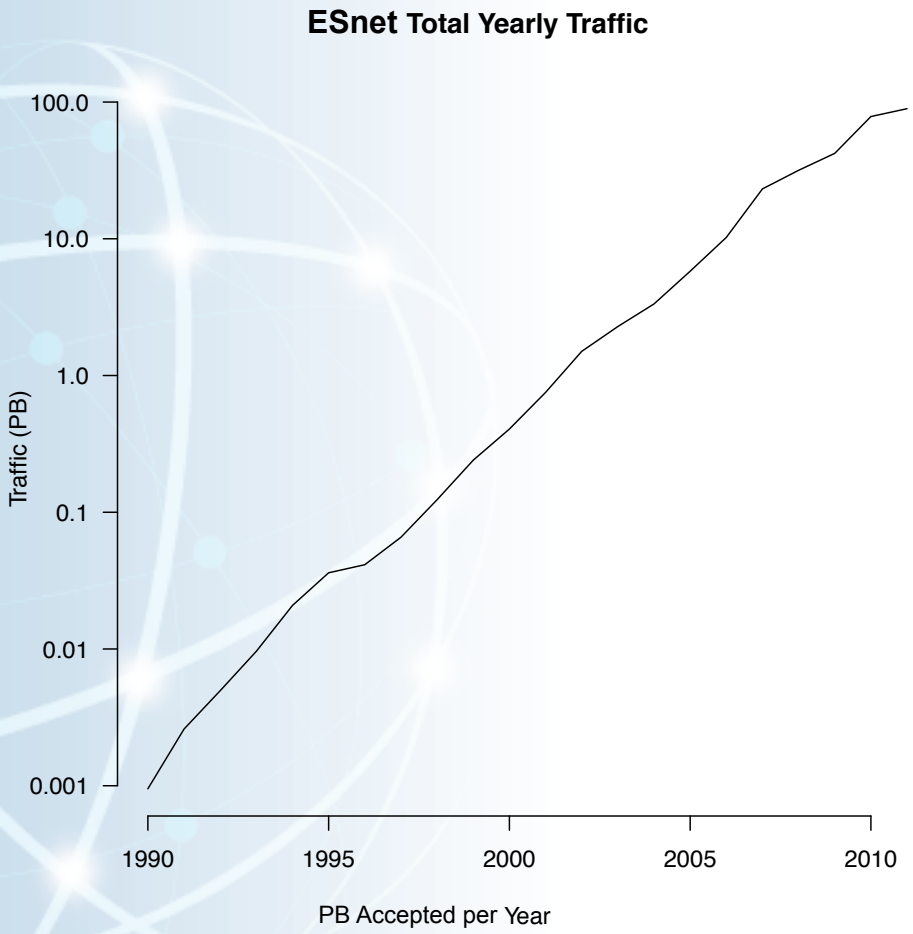


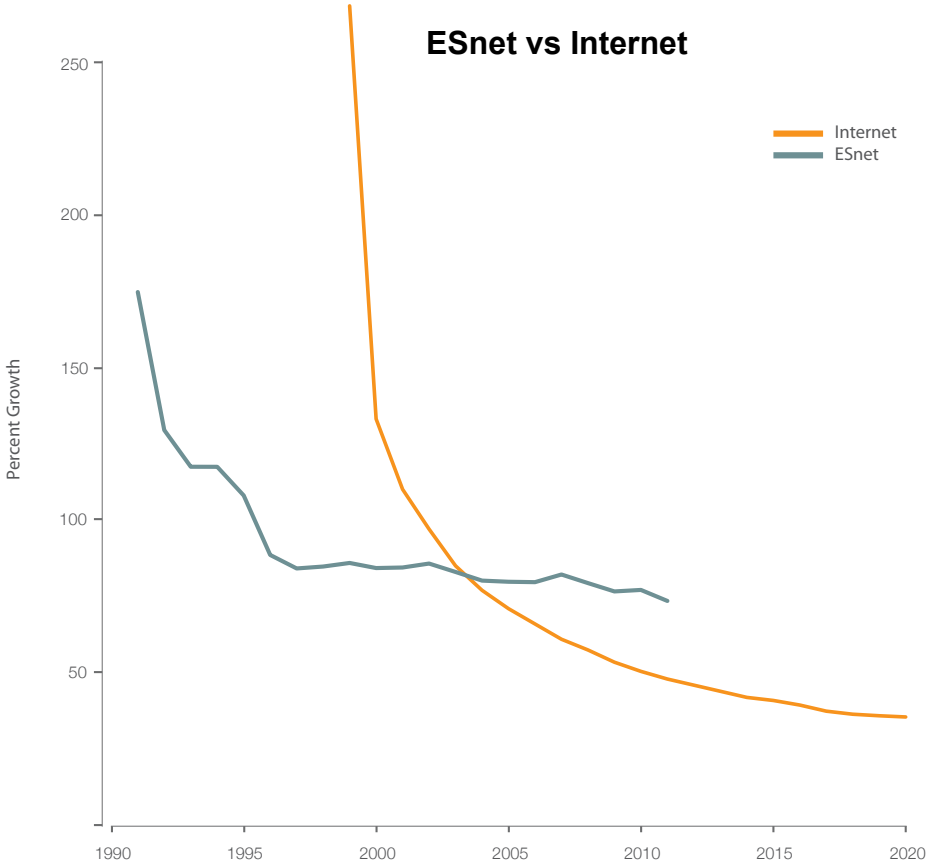
12,924 miles of Long-Haul Dark Fiber, ESnet5+



BayExpres Metro Fibers: 432 miles
ChiExpress Metro Fibers: 167 miles
NYExpress Metro Fibers: 6 miles

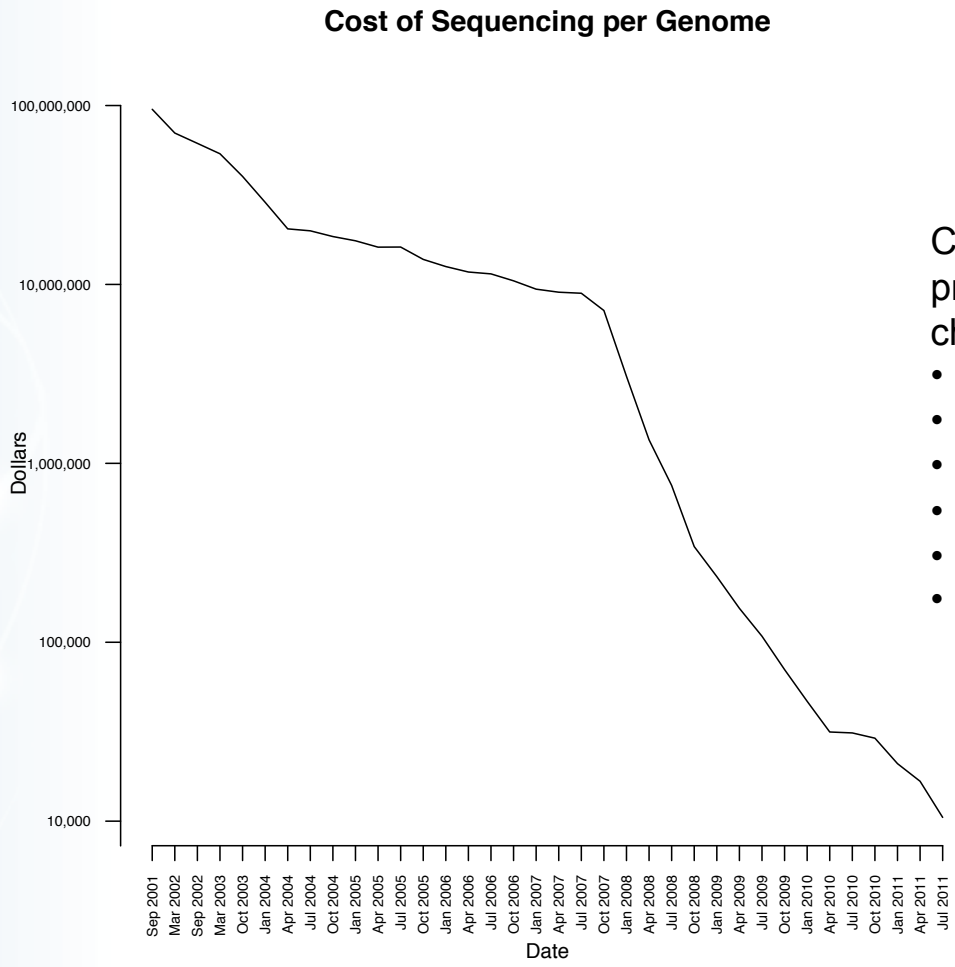
One Important **Driver for Innovation**





Compound Annual Growth Rates for ESnet [measured] and Internet core [estimated and predicted, see "Power Trends in Communication Networks", IEEE Journal of Selected Topics in Quantum Electronics, VOL. 17, NO. 2].

ESnet is Engineered for the Data ~~Explosion~~ ~~Flood~~ ~~Tsunami~~ Revolution

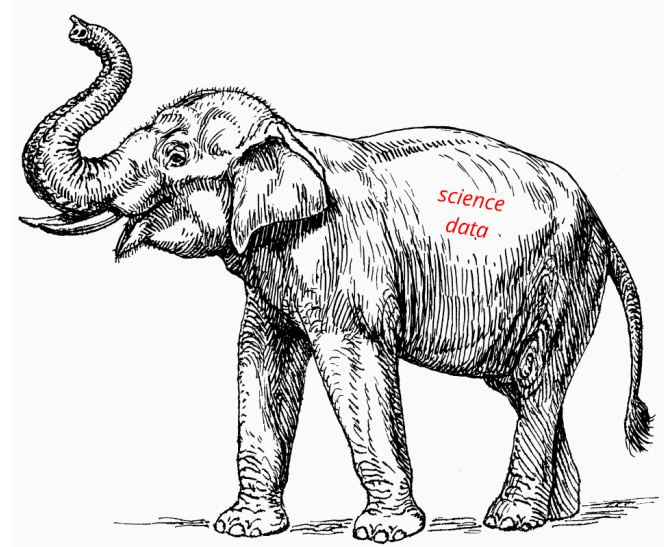


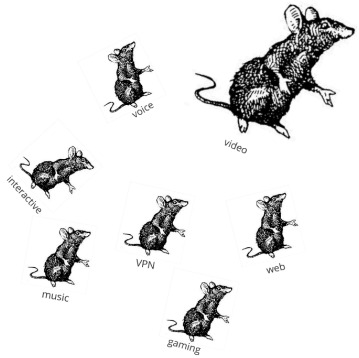
source: http://www.genome.gov/sequencingcosts/

Crushing data volumes producing **qualitative** changes in:

- instrument design
- workflows
- collaborations
- mission statements
- cyberinfrastructure
- discovery processes

Elephant (or Alpha) Flows





Flow-Type Zoology, by Dimension:

- size: elephant and mouse
- rate: cheetah and snail
- duration: tortoise and dragonfly
- burstiness: porcupine and stingray

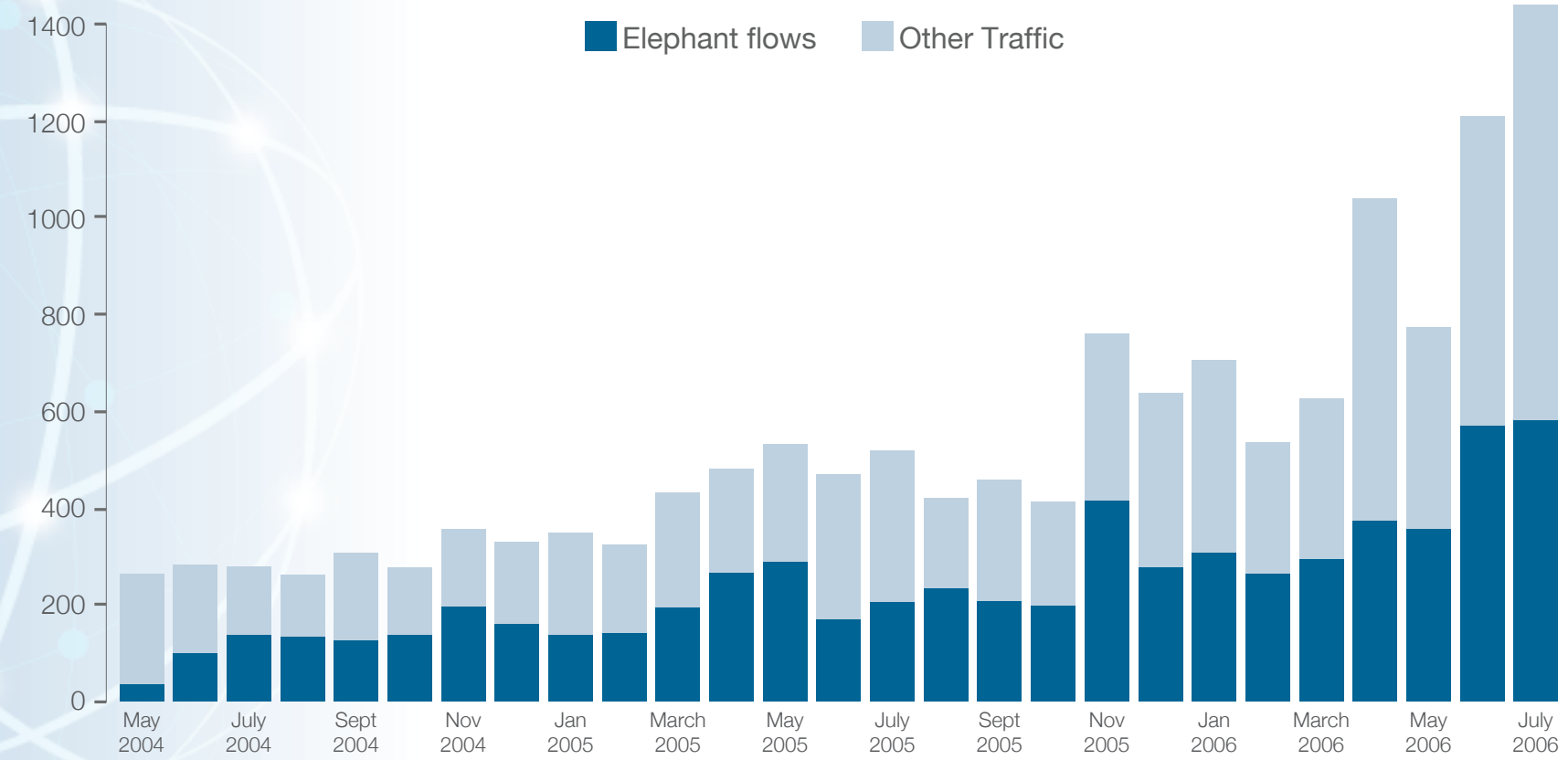
Kun-chan Lan and John Heidemann, "A Measurement Study of Correlations of Internet Flow Characteristics." *ACM Comput. Netw.* 50, 1 (January 2006), 46-6.

Or:

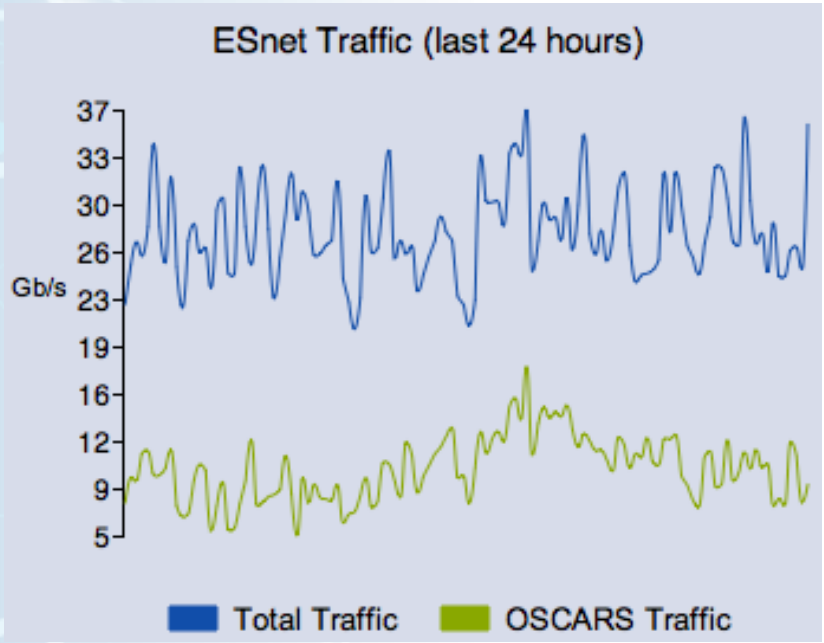
- Alpha flows (> H byes in < T sec)
- caused by transfers of large files over fast links

S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-Level Analysis and Modeling of Network Traffic," *ACM SIGCOMM Internet Measurement Workshop*, November 2001, 99–104.

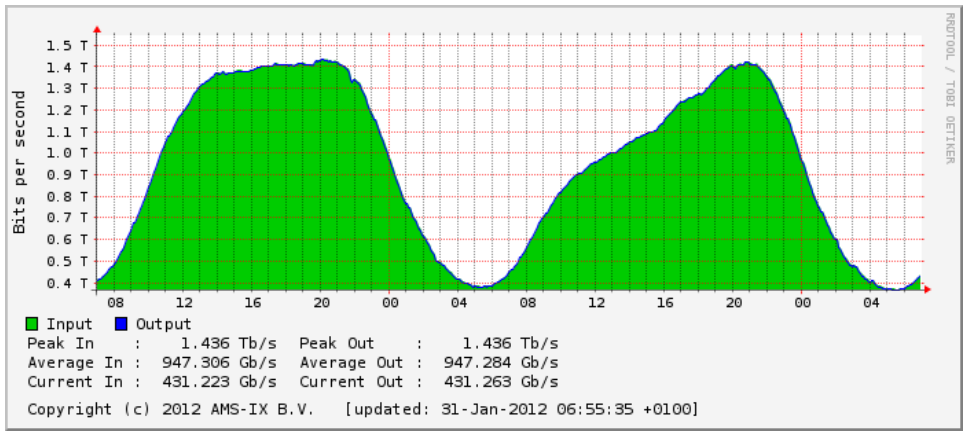
Science Traffic is Different



Science Traffic is Different



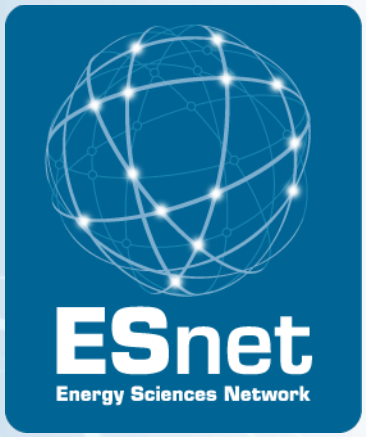
ESnet Accepted Traffic on January 28, 2012 (one day)
http://www.es.net



http://www.ams-ix.net/statistics/

How do we Know what our **Customers Need?**

- Each Program Office at Office of Science has a dedicated requirements workshop devoted to its needs
- Two requirements workshops per year, with attendees chosen by science programs
- Discussion centered on science case studies
- Requirements derived from case studies + discussions
- Reports contain requirements, case study text

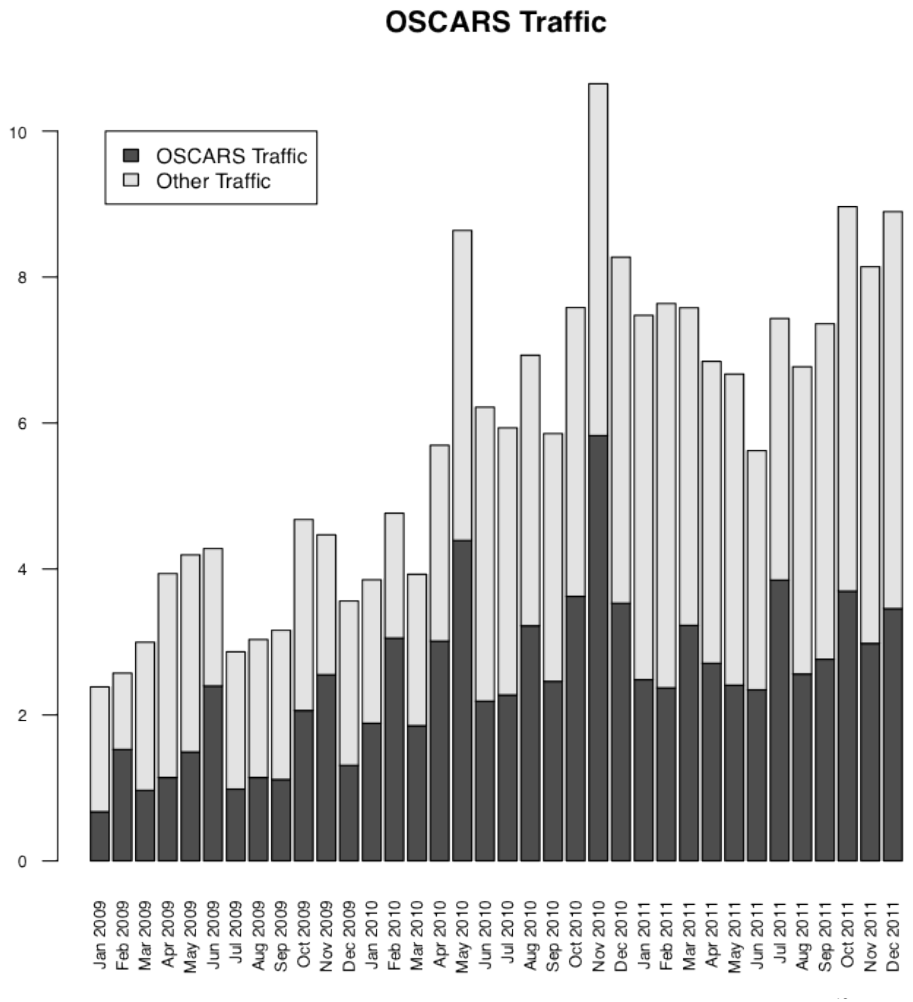


ESnet R&D

ESnet Research Success: OSCARS

OSCARS: an application and protocol suite for dynamic, inter-domain circuit management

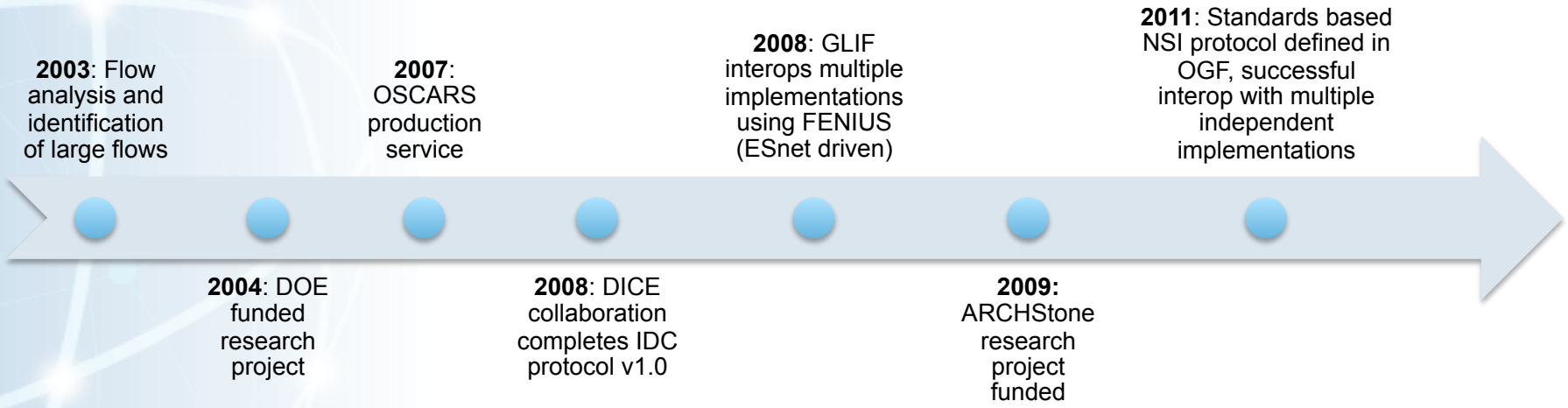
- *notable ESnet innovation*
- *used for service guarantees, traffic engineering, overlays*
- *carries almost ~50% of ESnet traffic (alpha flows)*
- *a customer-facing service: web interface for people, API for apps*
- *deployed in 21+ networks*



Production services
OSCARS

Research Motivation
- Network guarantees needed by service sensitive applications & large flows

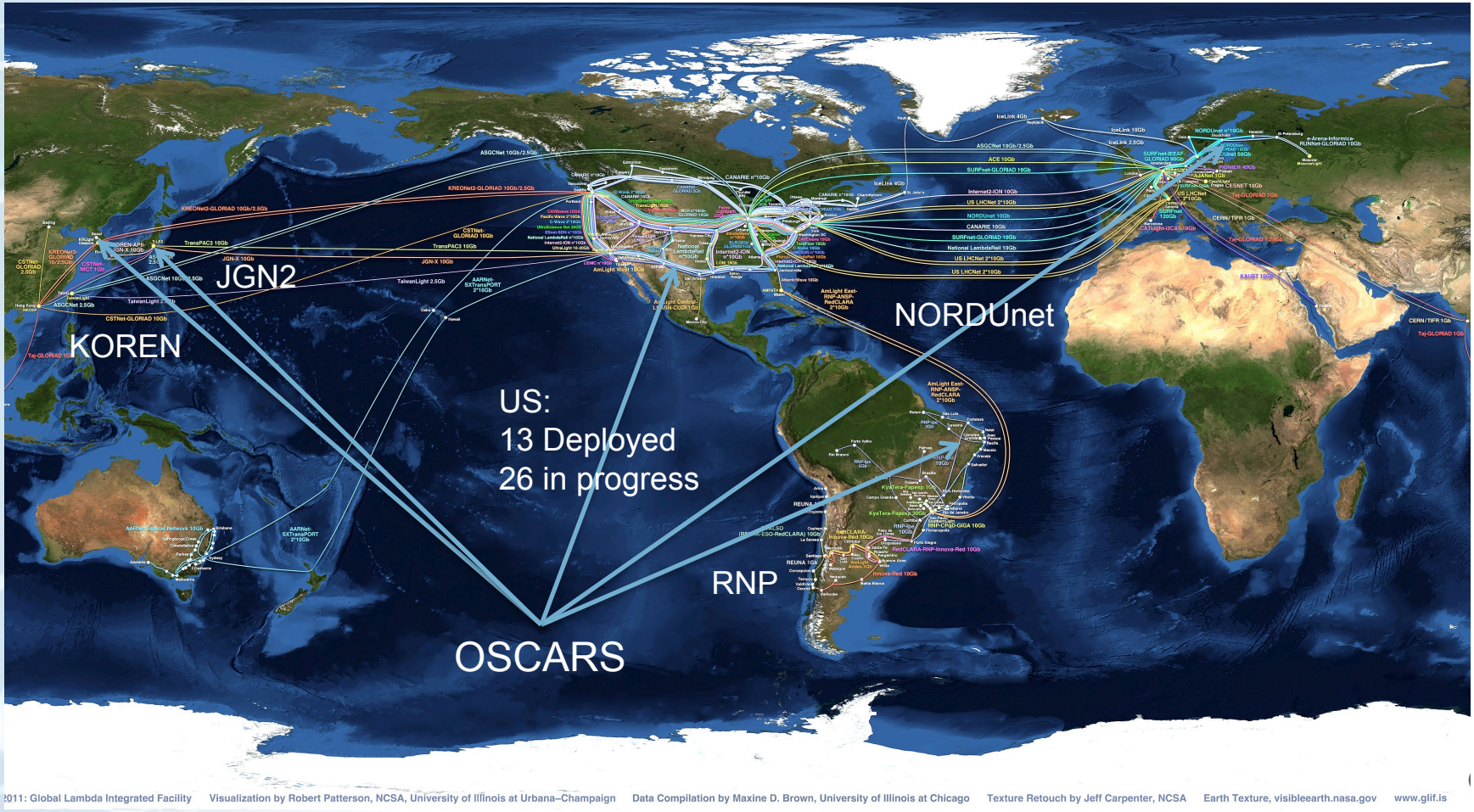
Objective
- Provide secure end-to-end guaranteed circuits across ESnet & multiple domains.



Impact
- Carries 50% of ESnet's 100+PB yearly data
- Deployed in about 21 networks
- Under evaluation by an additional 26 greenfield deployments

OSCARS Deployments Worldwide

WAN, Regional, Campus, Testbeds



ESnet Research Success: perfSONAR

Distributed platform and protocol suite for performance measurement, monitoring, and visualization.

Assures end-to-end health of global networks.

400+ nodes deployed worldwide.

Developed in a global collaboration of R&E networking organizations.

Helps us find and fix problems!

perfSONAR

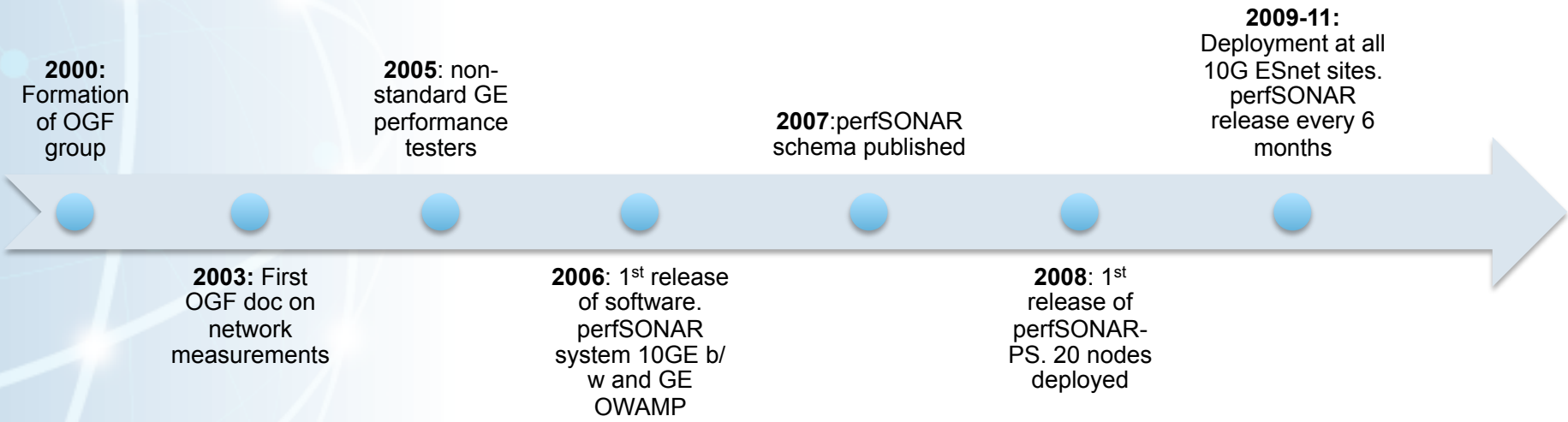
Network tools
perfSONAR

Research Motivation

- Certain performance problems only apparent to long-latency TCP flows, testing those requires source/sink in other domains

Objective

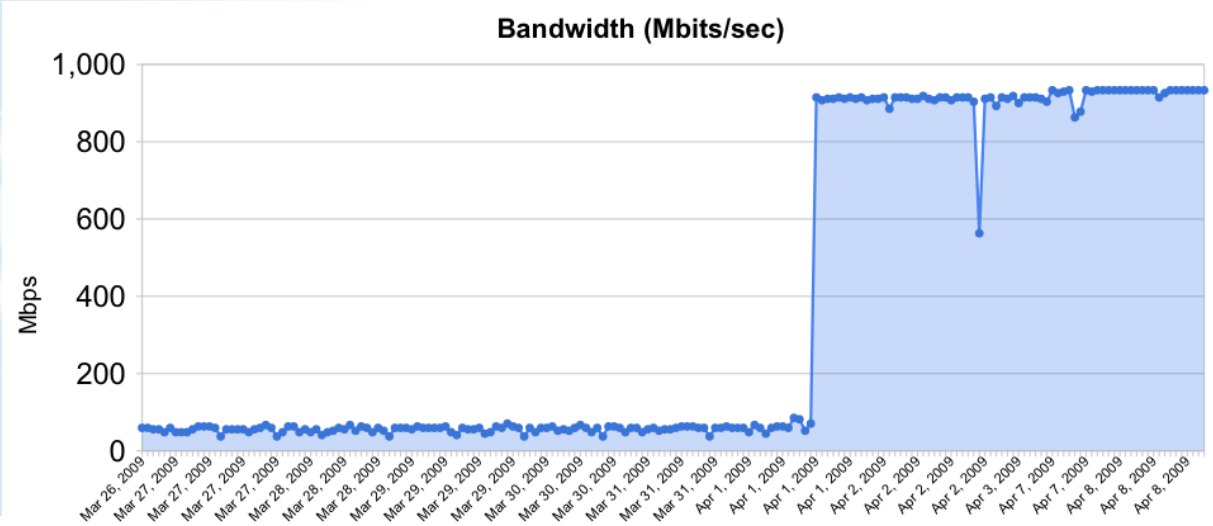
- Provide an infrastructure that identifies and isolates performance problems in a multi-domain environment.



Impact

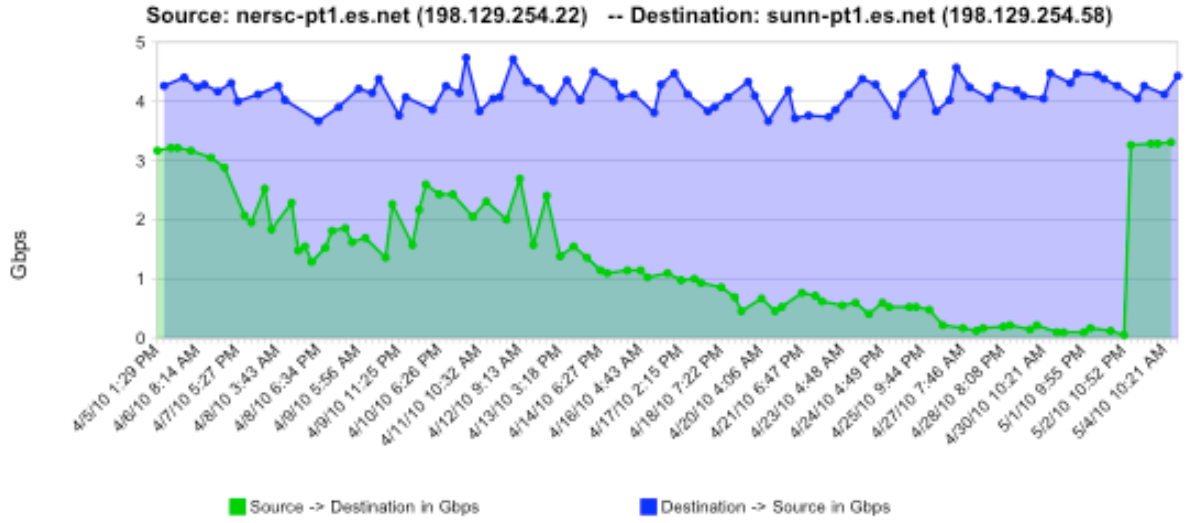
- Almost every ESnet site has fixed network performance problems
- LHC collaboration has deployed extensively
- 400+ perfSONAR hosts deployed on 100 networks and in 14 countries

perfSONAR in Use



Effect of full routing table, followed by reboot.

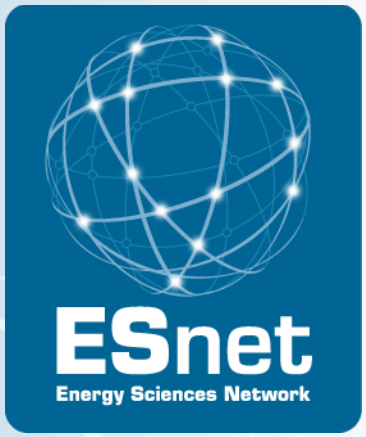
Effect of gradual optical line card failure.



Creating VO-Specific Dashboards from perfSONAR Data

Status of perfSONAR Throughput Matrix

-	0	1	2	3	4	5	6	7	8
0:atlas-npt2.bu.edu	-	OK OK	OK OK	OK OK	OK OK	OK OK	UNKNOWN OK	OK OK	OK OK
1:lhcmon.bnl.gov	OK OK	-	OK OK	OK OK	OK OK	OK OK	OK OK	OK UNKNOWN	OK OK
2:ps2.ochep.ou.edu	OK OK	OK OK	-	OK OK	OK OK	OK OK	OK UNKNOWN	OK OK	OK OK
3:psmsu02.aglt2.org	OK OK	OK OK	OK OK	-	OK OK	OK OK	UNKNOWN UNKNOWN	OK OK	OK OK
4:netmon2.atlas-swt2.org	OK UNKNOWN	UNKNOWN OK	OK OK	OK OK	-	OK UNKNOWN	OK UNKNOWN	OK OK	OK OK
5:iut2-net2.iu.edu	OK OK	OK OK	OK OK	OK OK	OK OK	-	OK OK	OK OK	OK OK
6:psnr-bw01.slac.stanford.edu	OK UNKNOWN	OK OK	UNKNOWN OK	UNKNOWN UNKNOWN	UNKNOWN UNKNOWN	OK OK	-	OK OK	UNKNOWN UNKNOWN
7:uct2-net2.uchicago.edu	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK	-	OK OK
8:psum02.aglt2.org	OK OK	OK OK	OK OK	OK OK	OK OK	OK OK	UNKNOWN UNKNOWN	OK OK	-



DOE's Advanced Network Initiative (ANI) Testbed Project Update

Testbed Value to Network Researchers

A realistic national-scale environment for innovative network research that would be impossible for most researchers to create in their labs

- Many types of network research require realistic high-latency environments

Maximum flexibility: researchers get “super-user” access to everything

- ability to install custom OS on hosts
- ability to do full routers/switch configuration
- ability to install custom router software

A controlled environment that supports reproducible results

Priority is given to research that aligns with ESnet strategic focus areas, such as:

High performance network protocols, data transfer middleware, Openflow + MPLS integration strategies, energy efficient networking

ESnet Research Testbeds

Fertile environment for researchers to test and prove concepts

Control Plane Testbed

- Long Island fiber, Infinera transport
- Juniper, NEC OpenFlow, Data transfer nodes, VMs

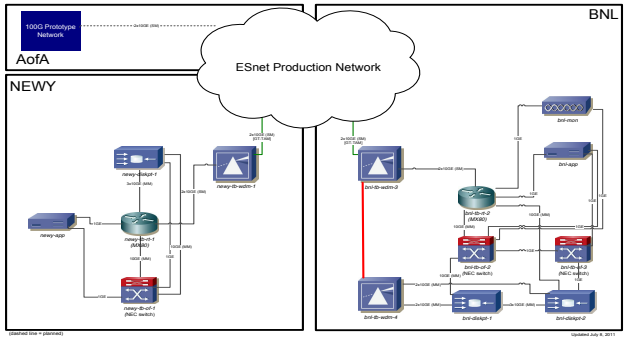
100G Testbed

- High-speed protocol research
- Routed network from Argonne to NERSC

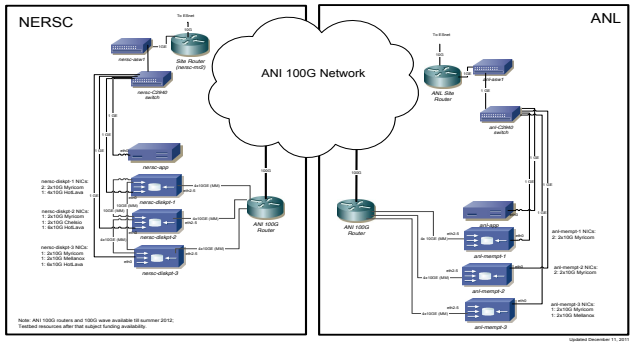
Dark Fiber Testbed

- Disruptive network approaches
- Quantum-key encryption

Long Island MAN (LIMAN) Testbed Architecture



ANI Middleware Testbed



LBNL Long Haul Dark Fiber Routes
12,924 miles

Community impact so far:
Peer-reviewed Testbed projects

- 25 projects accepted:
  - 17: testbed proposal review
  - 8: direct DOE/ASCR funding
  - 5 from Industry; 8 from DOE labs; 3 from NASA; 1 from DISA; 7 from Universities
- Strong diversity in research topics
- Four papers accepted or submitted for publishing, more to follow

Control Plane Testbed

Control Plane Testbed

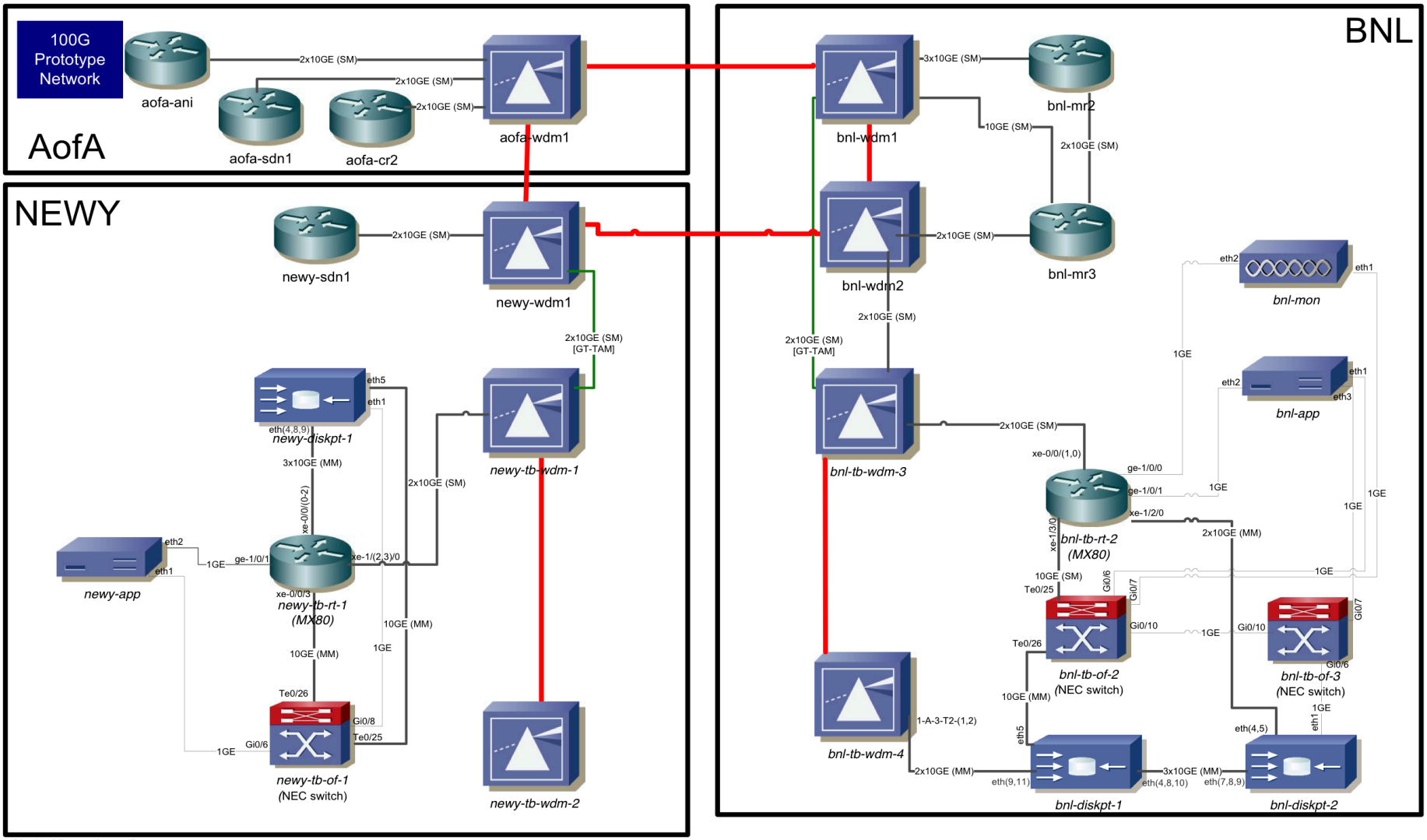
700 KM loop on Long Island Between Manhattan and Brookhaven National Lab

Infinera DWDM switches, NEC OpenFlow switches, Juniper router (soon with OpenFlow support), 6 hosts

Research Examples

- OpenFlow experiments
- Path Computation algorithms
- MPLS/OpenFlow Integration

Control Plane Testbed, Long Island, NY



(dashed line = planned)

Updated January 17, 2012

Notes:

-"App Host": can be used for researcher application, control plane control software, etc. Can support up to 8 simultaneous VMs

-"I/O Testers" are capable of 15 G disk-to-disk or 35G memory-to-memory

-Other infrastructure not shown: VPN Server, file server (NFS, webdav, svn, etc.)

100G Testbed

100G Testbed

100Gbps path from Oakland, CA (NERSC) to Argonne, IL (ANL)

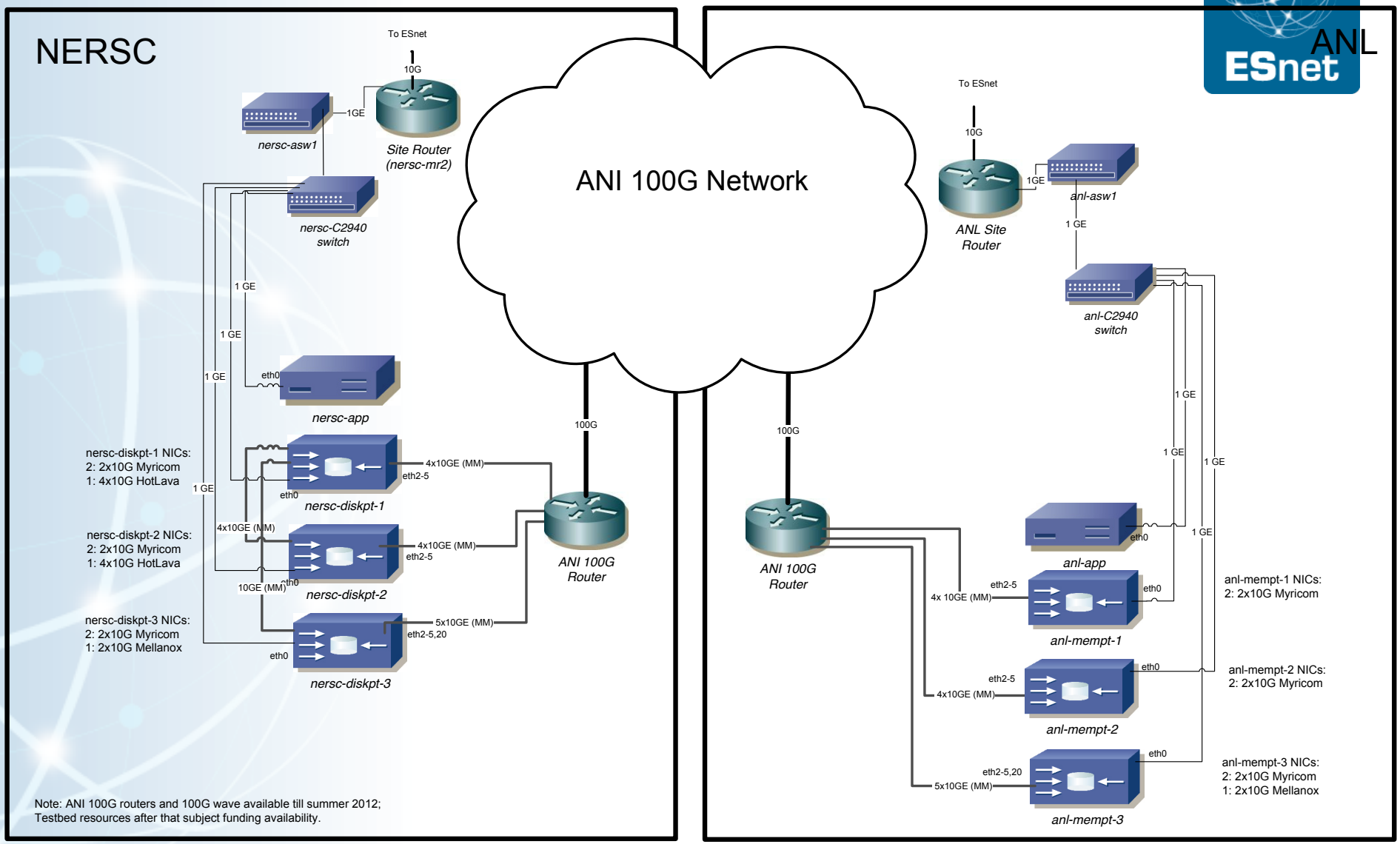
3 high-end hosts at each site, capable of >39 Gbps each

Research Examples

- 100G data transfer tools
  - FTP100, GridFTP, BestMan, Saratoga
- TCP alternatives
  - RDMA, RoCE
- TCP enhancements
  - PERT, RAPID TCP

Available as of Jan 3, 2012

ANI 100G Testbed



Related Resources

Magellan at NERSC

- currently 15 nodes with 1 10GE, GPFS backend
- soon: 8 nodes with 2x10GE, GPFS backend

Magellan at ANL

- went away in January

ORNL test cluster

- 12 nodes, each with flash-based scratch disk

FNAL test cluster

- in planning stage

Dark Fiber Testbed

Dark Fiber Testbed

Nationwide footprint (13,000 miles)

Researchers must pay to light the fiber, but no charge for access if results will be public

Research Examples

- Greentouch Energy experiments
- JPL quantum networking / quantum cryptography experiments

Dark Fiber Network

LBNL Long Haul Dark Fiber Routes
12,924 miles

BayExpres Metro Fibers: 432 miles
ChiExpress Metro Fibers: 167 miles
NYExpress Metro Fibers: 6 miles

Current ANI Testbed Research

25 projects have been accepted to the testbed so far:

- 8 via direct DOE/ASCR funding
- 17 via testbed proposal review process
- 5 from Industry; 8 from DOE labs; 3 from NASA, 1 from DISA, 7 from Universities

Wide range of projects:

- 9: high-speed middleware
- 3: OpenFlow
- 4: other network control plane
- 1: 100Gbps end host hardware
- 1: network flow classification
- 1: quantum communications
- 2: Wide Area RDMA
- 2: TCP congestion control
- 1: security
- 1: energy efficiency

http://www.es.net/RandD/advanced-networking-initiative/current-testbed-research/

Potential Breakthroughs from Current Testbed Research

Unconditionally secure continuous variable quantum communications

Ability to scale TCP well beyond 10Gbps

Alternative transport protocols that scale better than TCP (e.g.: RDMA)

Ability to easily create end-to-end circuits

Ability to scale data transfer tools/middleware to 100Gbps and beyond

100Gbps host interface

For more information see:

http://www.es.net/RandD/advanced-networking-initiative/current-testbed-research/

Results to Date: Publications Submitted

After only 7 months of being operational, there were already several publications submitted based on ESnet's Testbed results:

Energy Profiling of Network Elements for Rate Adaptation Technologies, M. Ricca, A. Francini, S. Fortune, T. Klein, submitted to the ACM/IEEE 3rd International Conference on Future Energy Systems (e-Energy 2012).

Traffic engineering in hybrid IP/optical circuit networks. Zhenzhen Yan, Chris Tracy, and Malathi Veeraraghavan, submitted to the IEEE 13th Conference on High Performance Switching and Routing (HPSR)

Middleware Support for RDMA-based Data Transfer in Cloud Computing, Yufei Ren, Tan Li, Dantong Yu, Shudong Jin, and Thomas Robertazzi, submitted to IEEE International Parallel & Distributed Processing Symposium (IPDPS)

Identifying gaps in Grid middleware on fast networks with the Advanced Network Initiative, G. Garzolglio and Dykstra, accepted at International Conference on Computing in High Energy and Nuclear Physics (CHEP 2012)

Results to Date: Press Releases

Joint Press Release by Orange Silicon Valley, Bay Microsystems and ESnet on the first wide area 40G RDMA results (http://goo.gl/57gSJ)

- Showed that RDMA moves data at up to 96 percent of the peak capacity of the network
  - Much more efficient than TCP
- Replicated at SC11 over a 10,000KM path (http://goo.gl/ZE0w8)

Greentouch press release on plans to use ESnet dark fiber testbed for energy experiments: http://goo.gl/OB6nD

Results to Date: Demonstrations

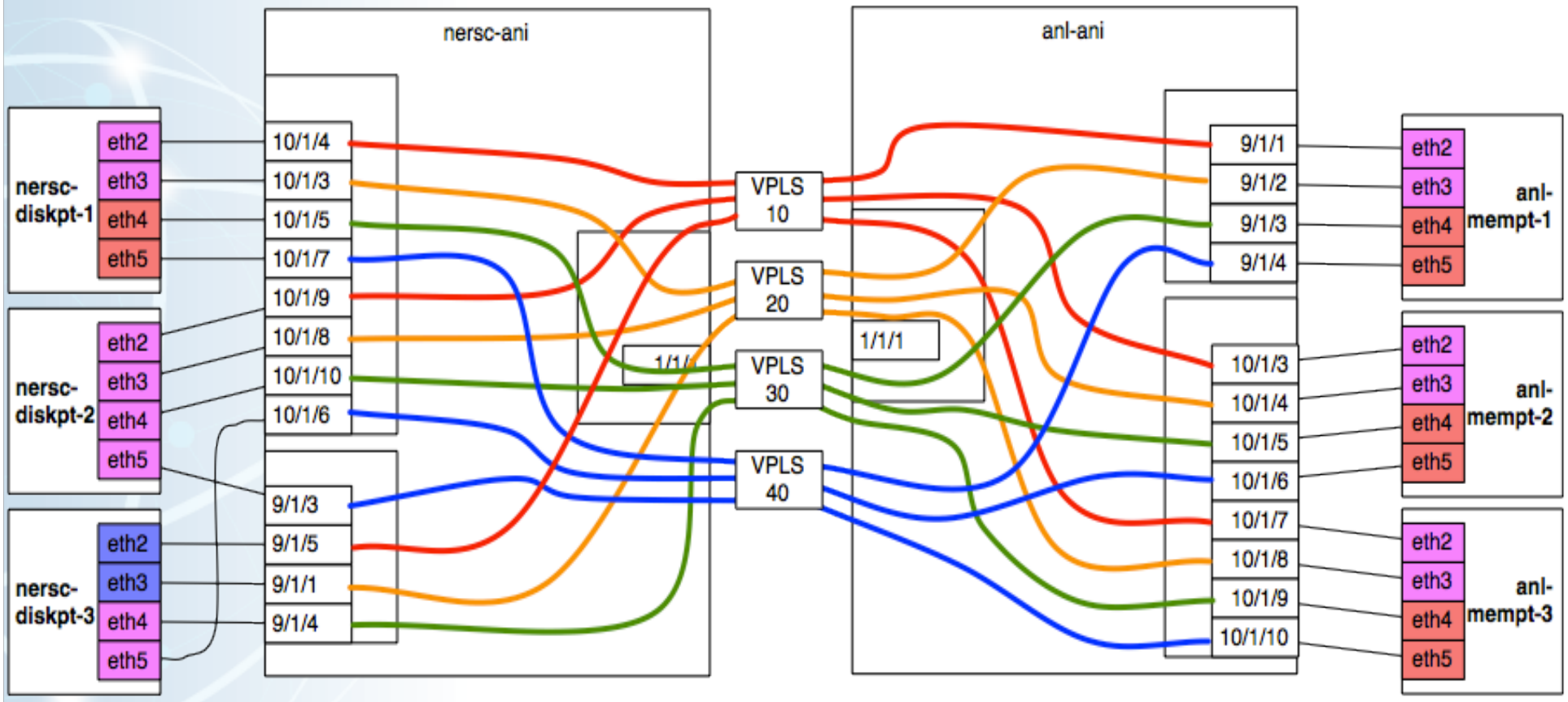
- ESnet Demonstration of OSCARS/OpenFlow integration on the ANI testbed at the Open Networking Summit held at Stanford.
  - http://goo.gl/4FFBg

- ESnet Demonstration of 9.9Gbps RDMA transfers from the ANI Testbed at BNL to Seattle at SC11

Results to Date: Industry Involvement

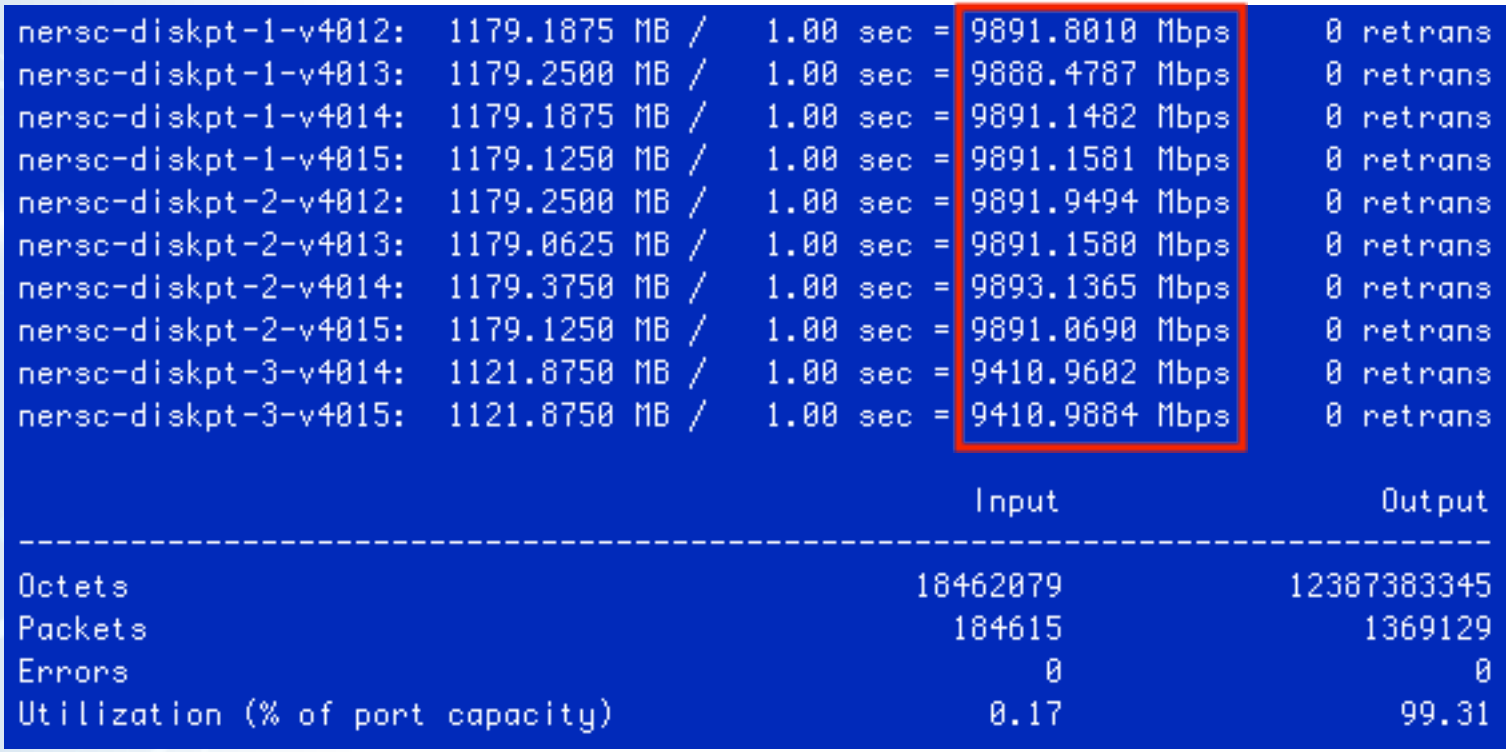
- Loaner hardware for 40GE testing
  - Infinera, Alcatel-Lucent (ALU), Bay Microsystems, Mellanox

100G Testbed Performance



48.6ms RTT, 97.9Gbps aggregate TCP throughput* with 10 TCP streams

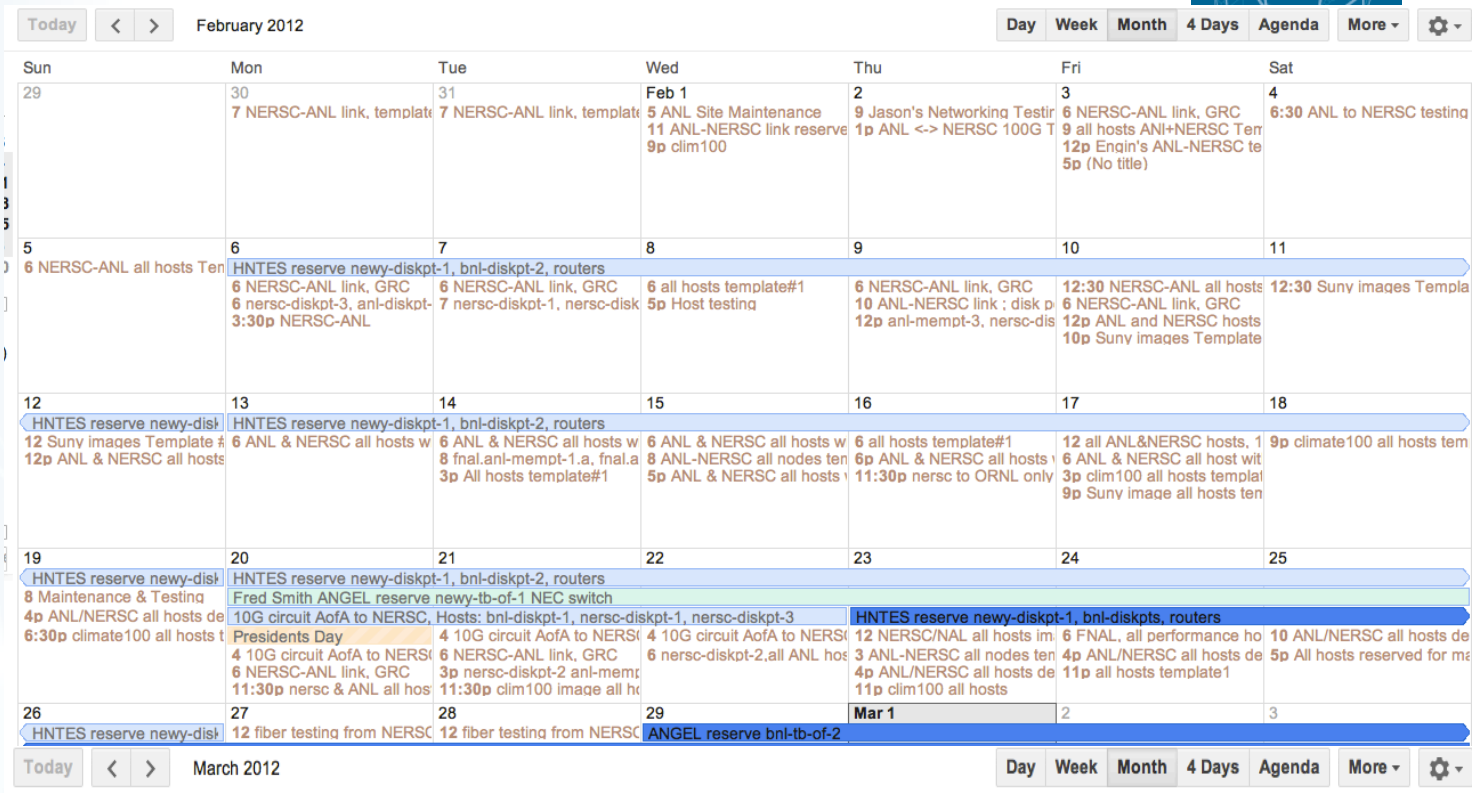
Per-Stream Results



Testbed Usage: Fully booked 24x7!

Feb 2012

March 2012



Testbed Access

Proposal process to gain access described at:

https://sites.google.com/a/lbl.gov/ani-testbed/

Testbed is available to anyone:

- DOE researchers
- Other government agencies
- Industry

Must submit a short proposal to the testbed review committee

- Committee is made up of members from the R&E community and industry

Goal is to accept roughly five proposals every 6 month review cycle

- Next round of proposals is due April 1, 2012

Future Testbed Challenges

Long term operational support of the Testbed

- Current testbed funding will run out in August of this year
  - Currently funded at 2FTE for testbed support
    - This is essential for continued success

Hope to secure funding to ensure persistent/indefinite availability

- 100G wave from NERSC to ANL will transition to production use starting this summer, so need to identify funds to light a 2nd wave for testbed use

Dark Fiber will be available to the testbed until it is needed for production use

- For some segments this may not be for several years.

More Information

http://sites.google.com/a/lbl.gov/ani-testbed/

email: ani-testbed-proposal@es.net

BLTierney@es.net